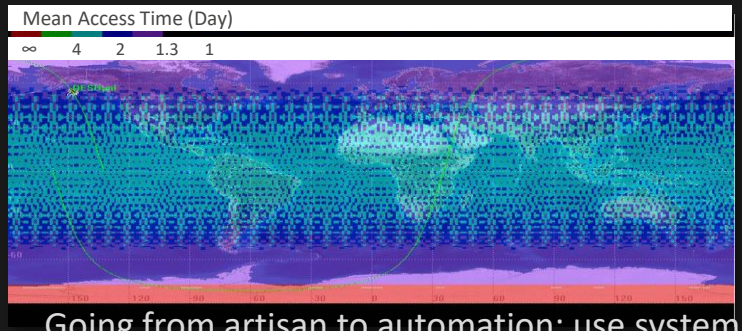
NASA JPL: Auto Scaling Advantage

Advanced Rapid Imaging and Analysis

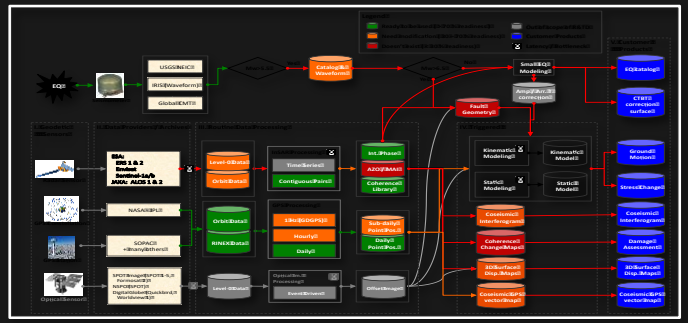
Hook Hua, Jet Propulsion Laboratory, California Institute of Technology

Jet Propulsion Laboratory
California Institute of Technology

Challenge: Automated and rapid remote sensing for urgent disaster response

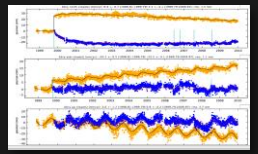


Automated data system are required to analyze large quantities of data from NASA NISAR, other satellite missions, and rapidly expanding GPS networks

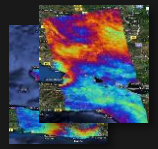


Going from artisan to automation: use system engineering approach to translate specialized data analysis into operational capability

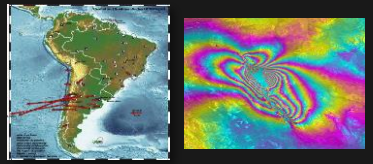
Demonstrate response to hazards with standardized set of data products for decision and policy makers



Temporal records of ground deformation

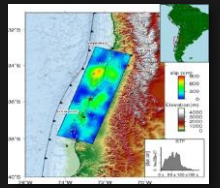


Spatial maps of ground deformation



Coseismic ground deformation

Coseismic damage



Earthquake Models

AWS re:Invent

Advanced Rapid Imaging and Analysis (ARIA)

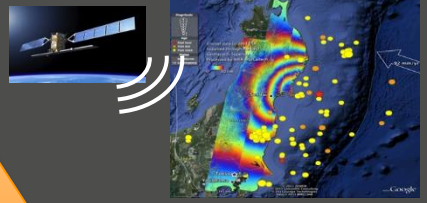
Radar Sensors

GPS Networks

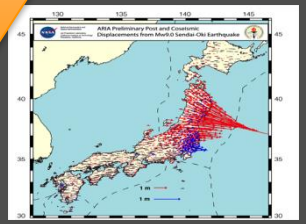
Seismic Networks

Optical Sensors

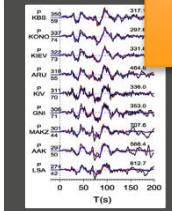
Automated data collection & processing



Radar



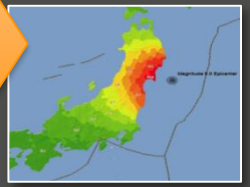
GPS



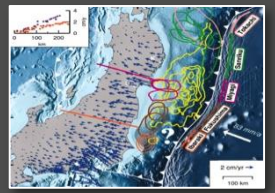
Seismology

Monitoring & near real-time assessment

Building damage and inundation



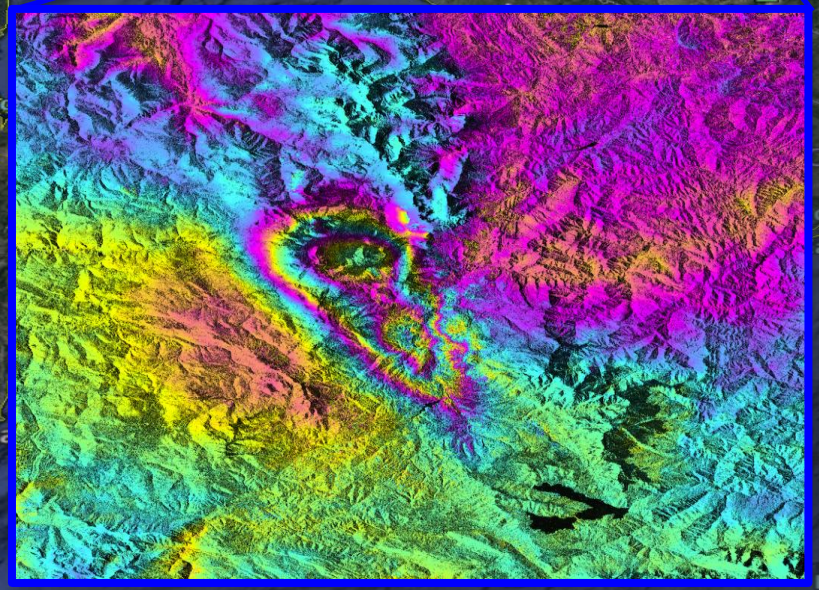
Permanent ground deformation



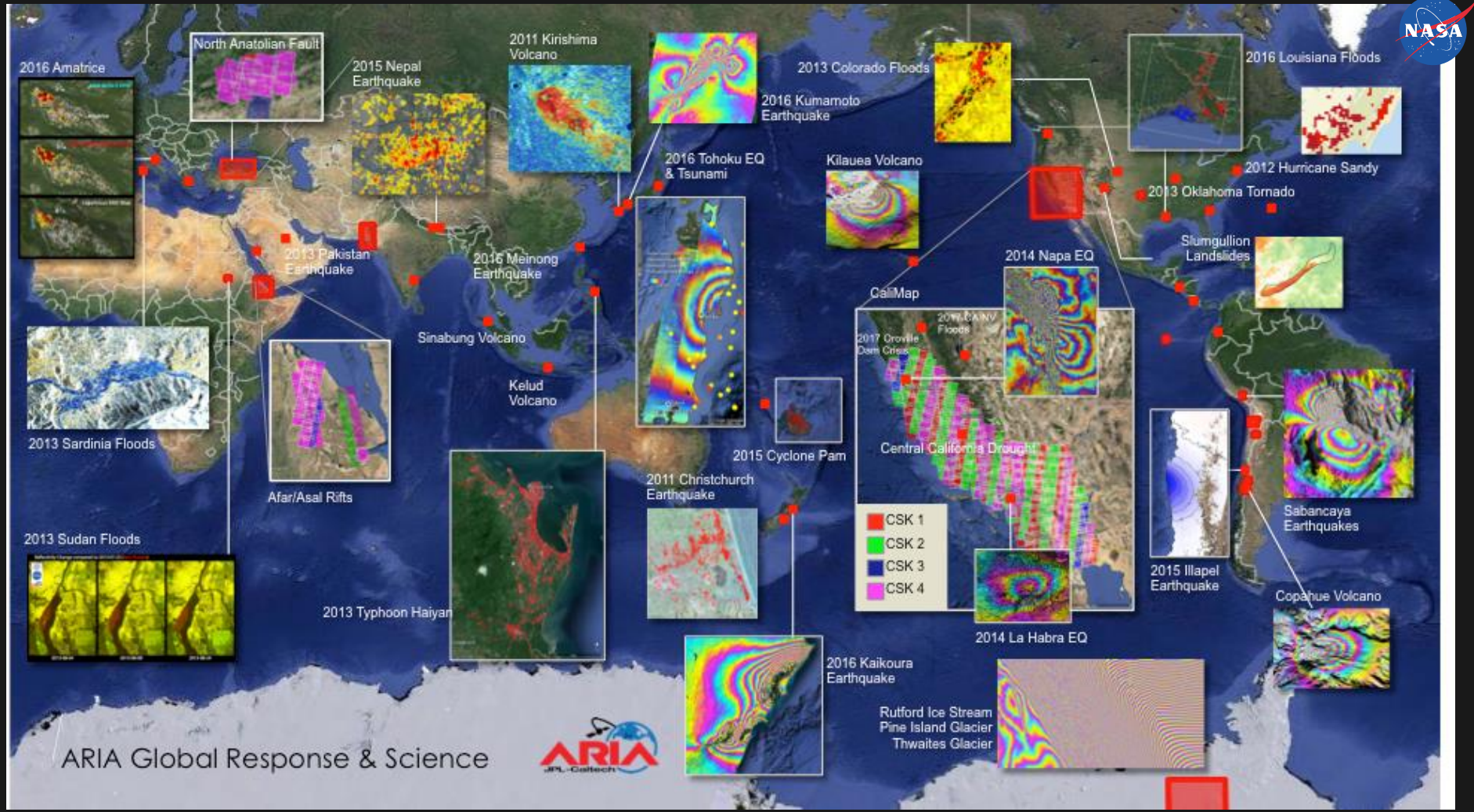
High-resolution hazard assessment from fault models

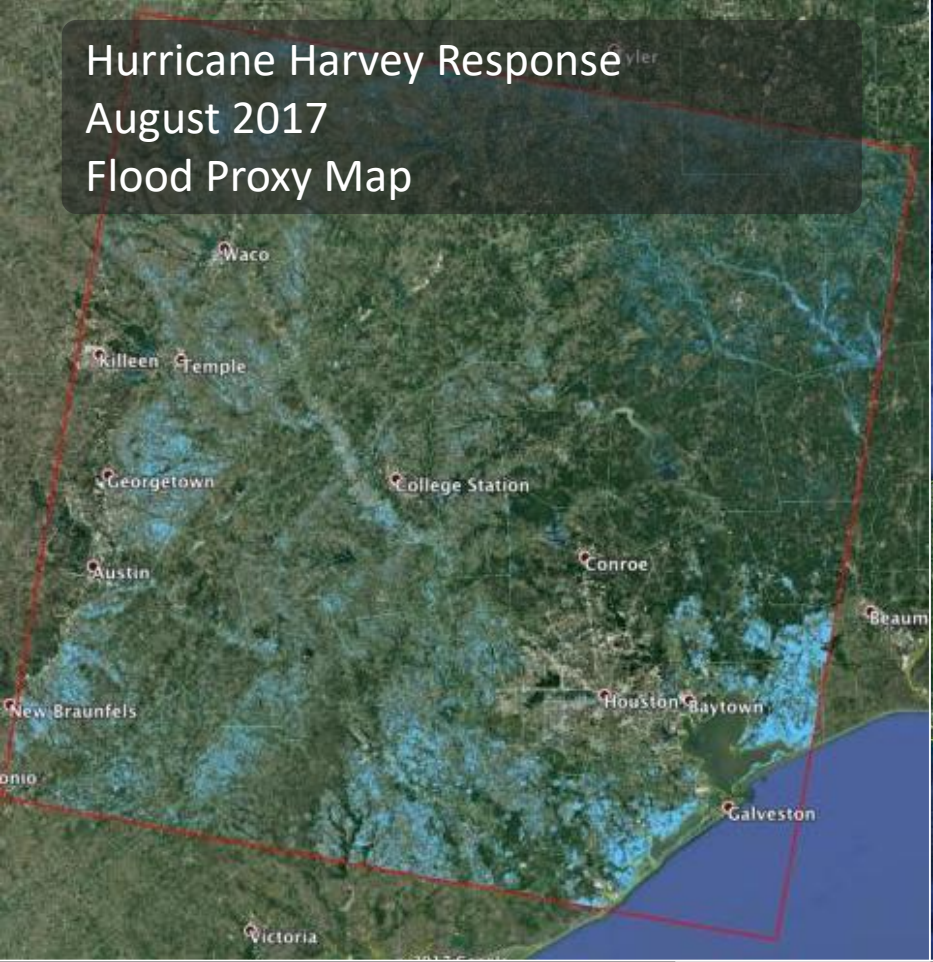
Examples from the 2011 M9.0 Tohoku-Oki (Japan) earthquake

Amatrice, Italy earthquake (August 23, 2016) Automated Urgent Response Interferogram



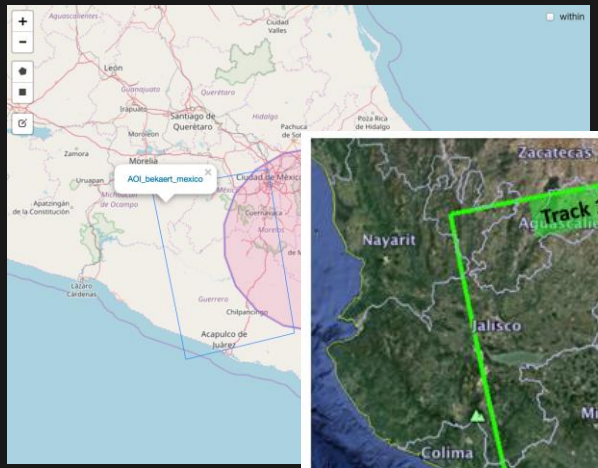
Urgent response processing of ESA's Sentinel-1A data to interferograms were automatically processed—all in AWS

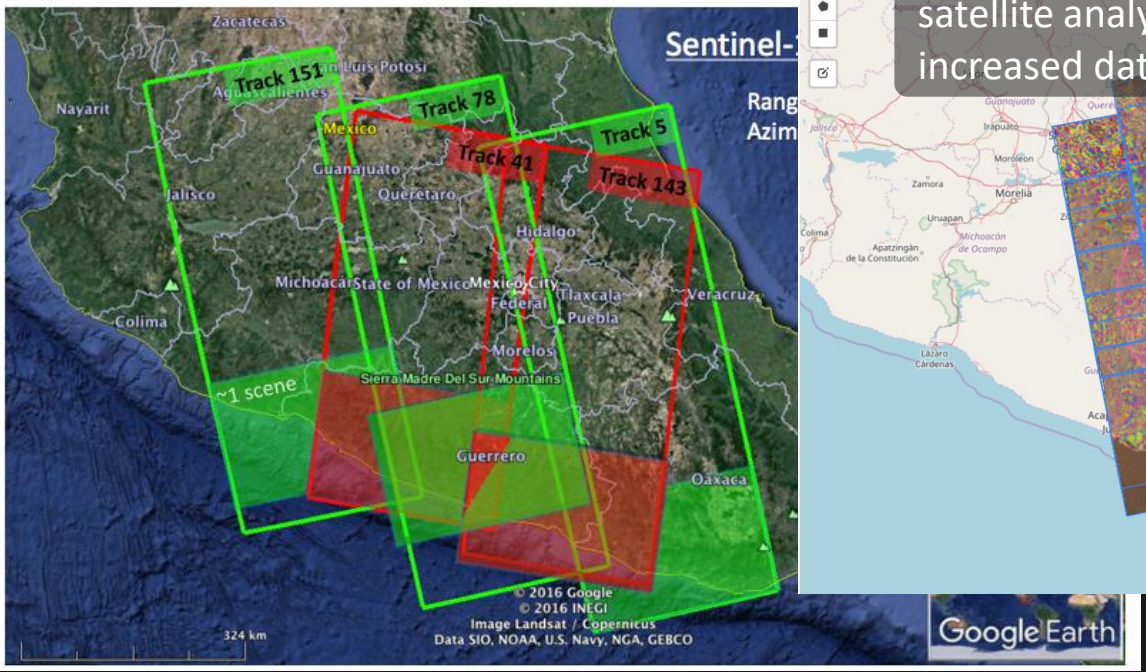

2016 Amatrice
North Anatolian Fault
2015 Nepal Earthquake
2011 Kirishima Volcano
2016 Kumamoto Earthquake
2013 Colorado Floods
2016 Louisiana Floods
2012 Hurricane Sandy
2013 Oklahoma Tornado
2016 Tohoku EQ & Tsunami
Kilauea Volcano
2013 Pakistan Earthquake
2016 Meinong Earthquake
Slumgullion Landslides
2014 Napa EQ
CalMap
2017 Oroville Dam Crisis
2017 UCANV Floods
Sinabung Volcano
Kelud Volcano
2013 Sardinia Floods
Afar/Asal Rifts
2015 Cyclone Pam
Central California Drought
2011 Christchurch Earthquake
CSK 1
CSK 2
CSK 3
CSK 4
Sabancaya Earthquakes
2015 Illapel Earthquake
2013 Sudan Floods
2013 Typhoon Haiyan
2014 La Habra EQ
Copahue Volcano
2016 Kaikoura Earthquake
Rutford Ice Stream
Pine Island Glacier
Thwaites Glacier
ARIA Global Response & Science
ARIA
JPL - Caltech

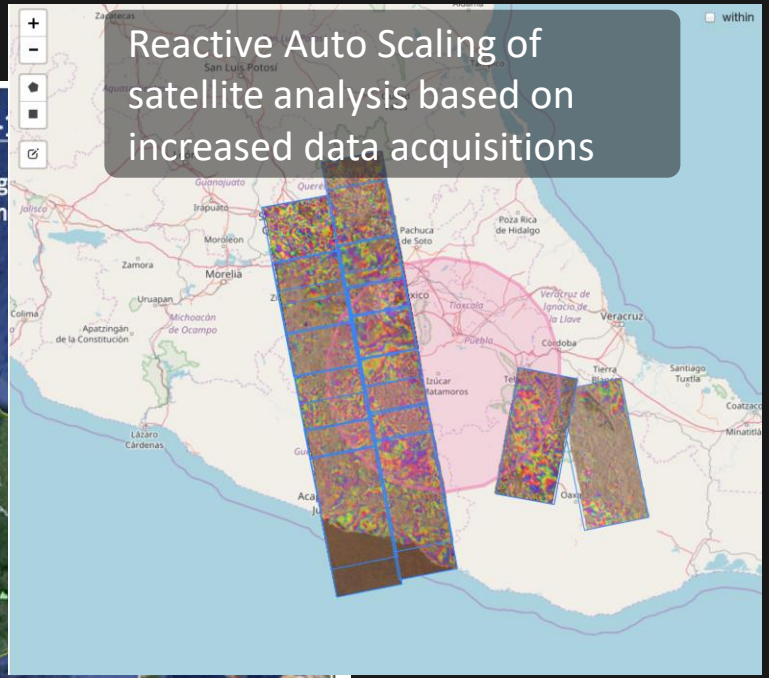


Urgent Response Analysis in AWS Cloud

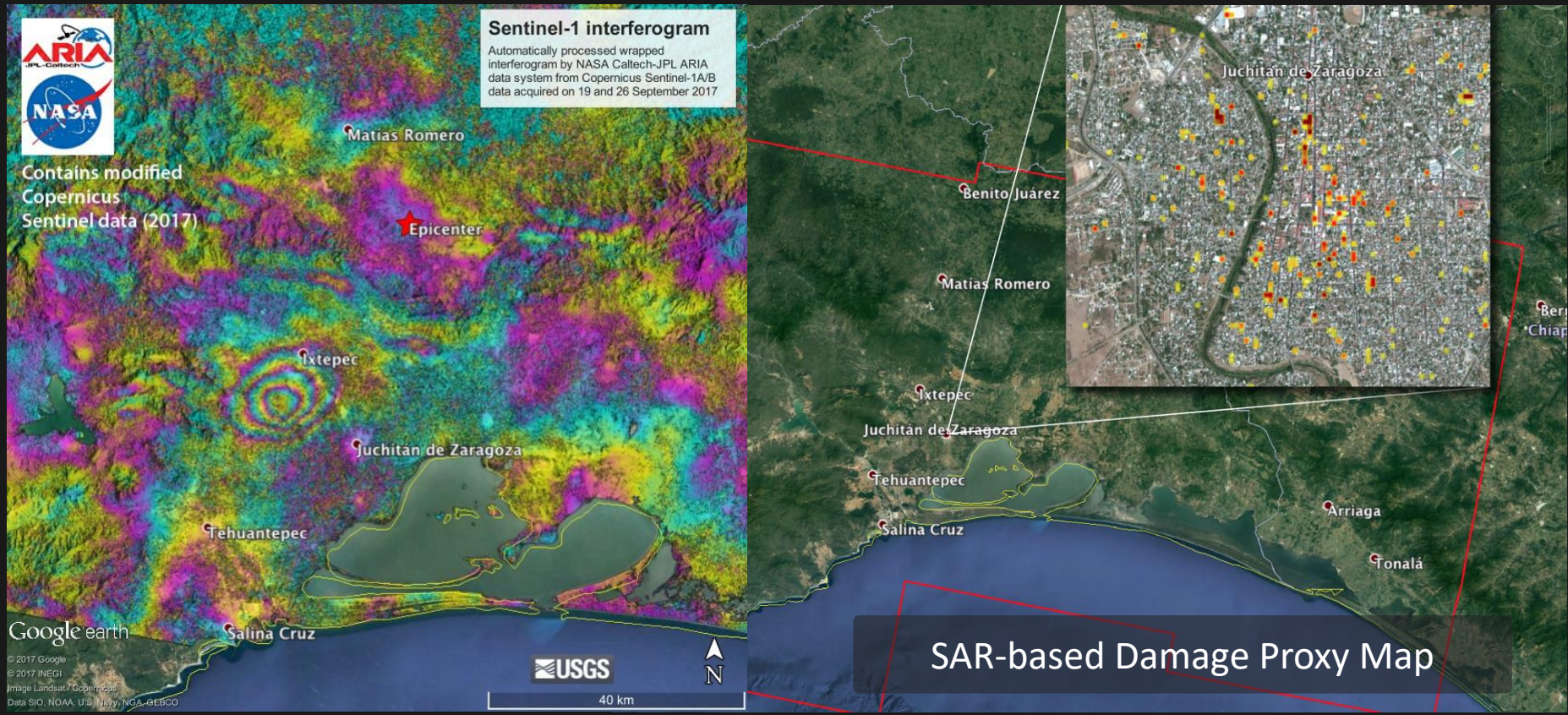
M7.1 Earthquake near Puebla, Mexico (September 9, 2017)







Urgent Response Analysis in AWS Cloud

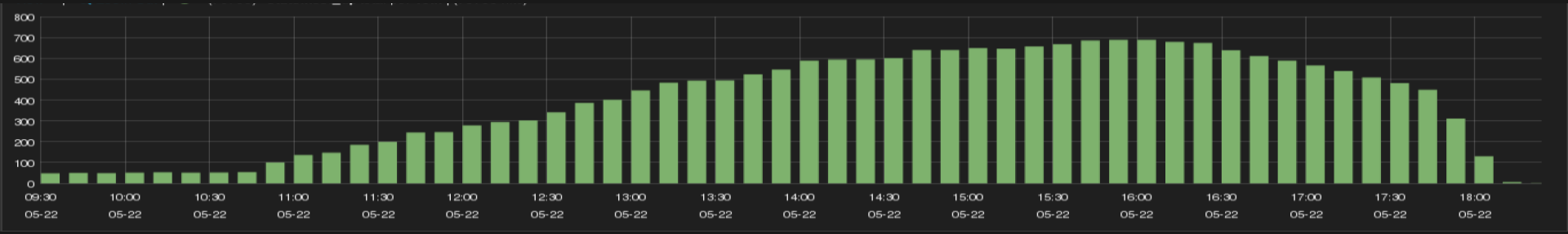


Dynamic Scaling in Earth Science Data System

The size of the science data system compute nodes can automatically grow/shrink based on processing demand

Auto Scaling group policies
Target tracking scaling policies

Auto Scaling enabling runs of over 100,000 ***vCPUs***

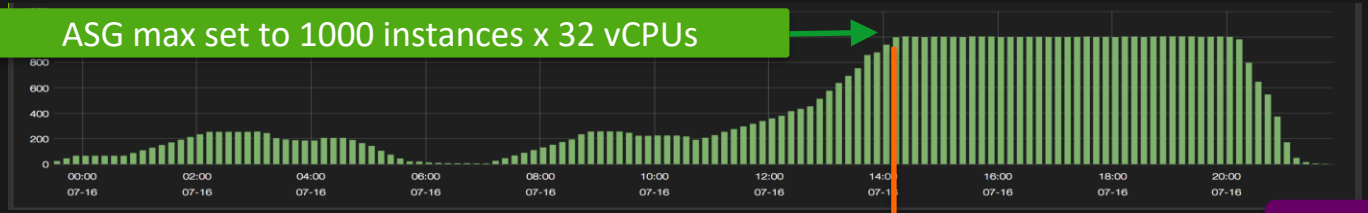


Earth Science Data System in AWS

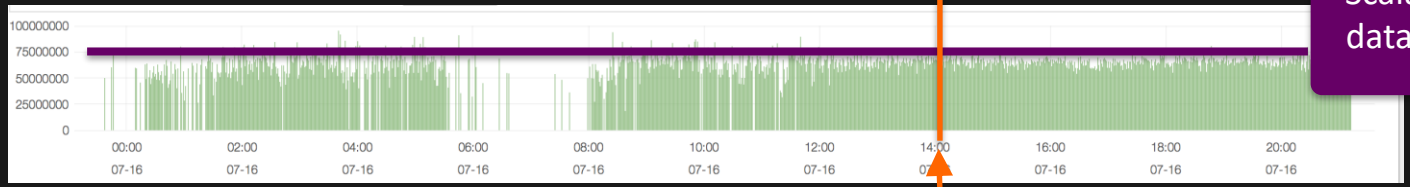
NASA OCO-2 L2 Full Physics processing operational in AWS

- Processing of L2 full physics data products in Amazon cloud across multiple regions
- Scaled up **thousands** of compute nodes
- Demonstrated capability of higher internal data throughput rates than NISAR needs

Number of compute nodes over time



Per node transfer rate over time



Scalable internal data throughput

@ 32,000 full-physics processing on 1,000 nodes

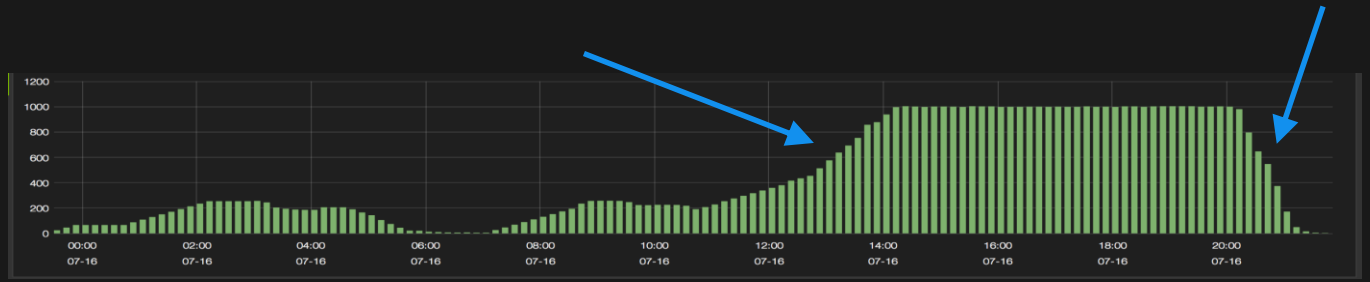
Considerations for Scaling In/Out Events

Scaling up (scale out)

- **Target tracking scaling policies**
- Scaling up in batches + rest periods

Scaling down (scale in)

- What policy to set to scale down? CPU/network utilization
- Potential **stateful** domain knowledge only known within the instances
- **Instance protection**

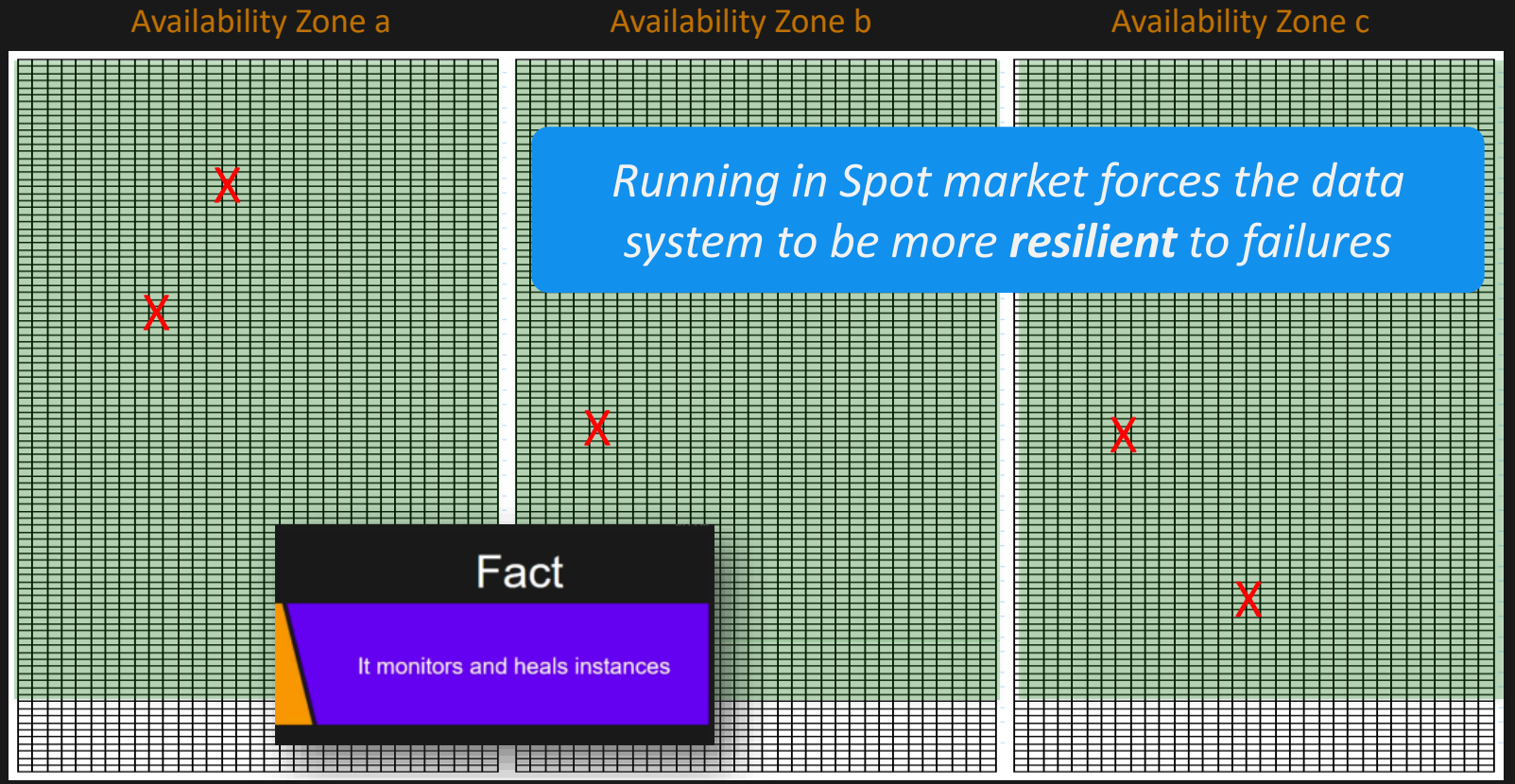


Auto Scaling and the Amazon EC2 Spot market

					US-West-2 (Oregon)							
					Hourly Costs				Per vCPU Costs			
instance	vCPU	memory	memory-cpu ratio	disks	on-demand ($/hr)	reserved 1-yr upfront ($/hr)	reserved 3-yr upfront ($/hr)	spot linux ($/hr)	on-demand ($/cpu/hr)	reserved 1-yr upfront ($/cpu/hr)	reserved 3-yr upfront ($/cpu/hr)	spot linux ($/cpu/hr)
m2.4xlarge	8	68.4	8.55	2 x 840	$1.0780	$0.4087	$0.244	$0.1000	$0.1348	$0.0511	$0.030	$0.0125
cc2.8xlarge	32	60.5	1.89	4 x 840	$2.0000	$0.9131	$0.613	$0.2705	$0.0625	$0.0285	$0.019	$0.0085
m3.2xlarge	8	30.0	3.75	SSD 2 x 80	$0.6160	$0.3750	$0.230	$0.0700	$0.0770	$0.0469	$0.028	$0.0088
c3.8xlarge	32	60.0	1.88	SSD 2 x 320	$1.6800	$0.9920	$0.628	$2.4001	$0.0525	$0.0310	$0.019	$0.0750
r3.8xlarge	32	244.0	7.63	SSD 2 x 320	$2.8000	$1.4860	$0.982	$2.8000	$0.0875	$0.0464	$0.030	$0.0875
c3.xlarge	4	7.5	1.88	SSD 2 x 40	$0.2310	$0.1370	$0.087	$0.0353	$0.0578	$0.0343	$0.021	$0.0088

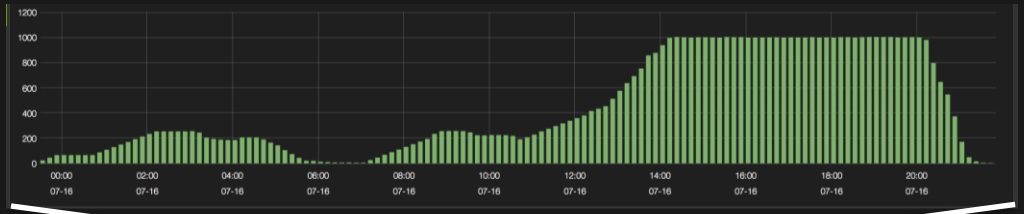
- Auto Scaling works well with Spot Instances
- Major cost savings (**75%–90%** savings over on-demand)…if can use **Spot Instances**
- Compute instances terminated if market prices exceed your bid threshold

Fleet Management for High Resiliency

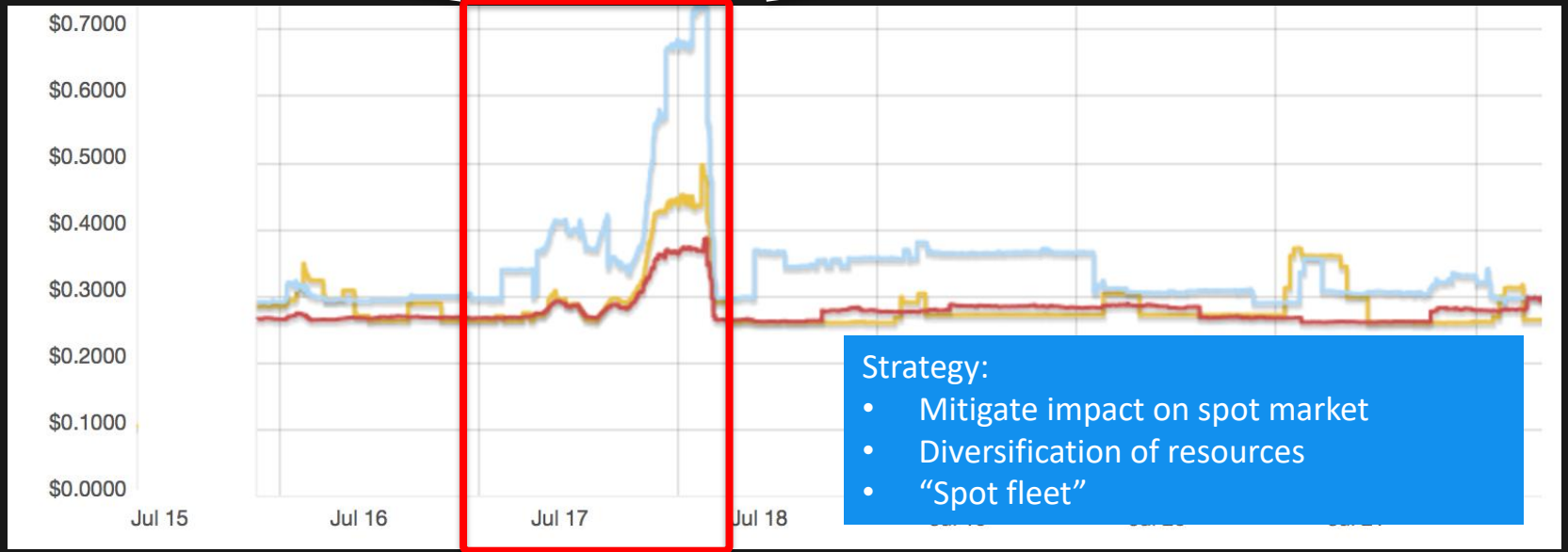


Compute fleet instances

Auto Scaling and the “Market Maker”



This OCO-2 data production run of 1000 x 32vCPUs affected the market prices



Strategy:

- Mitigate impact on spot market
- Diversification of resources
- “Spot fleet”

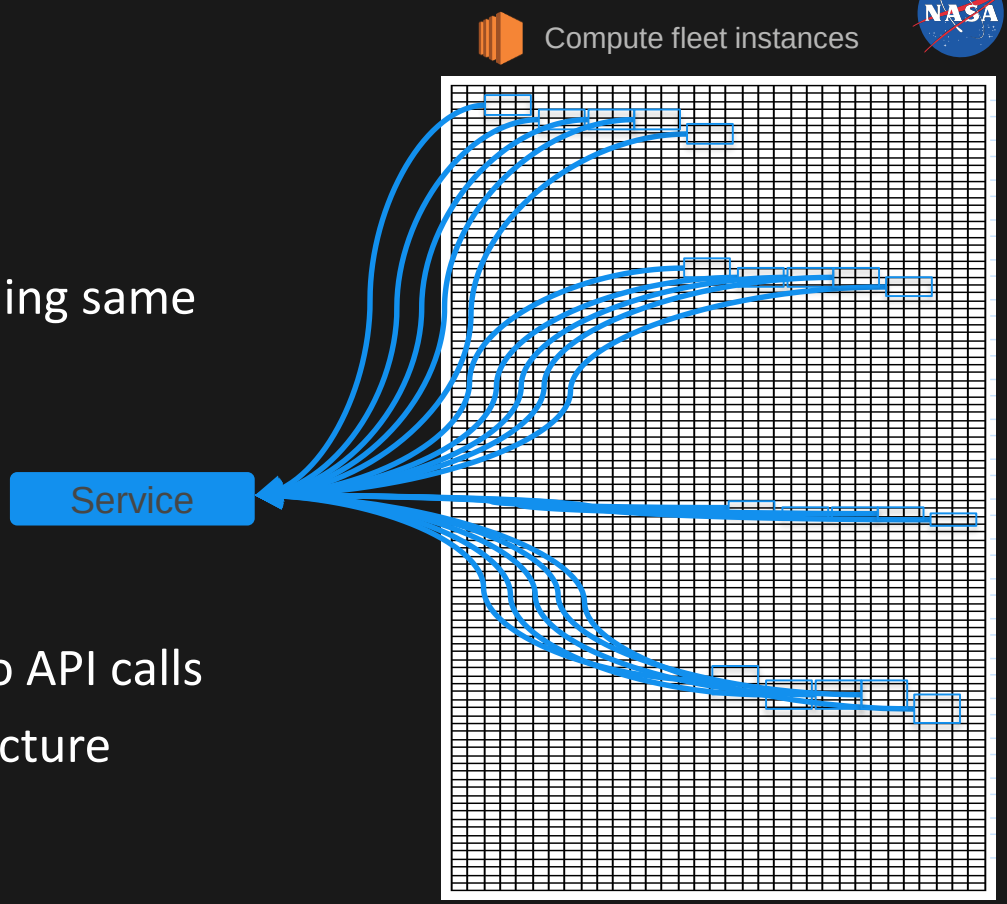
“Thundering Herd”

Fleet of ASG compute instances calling same services at same time

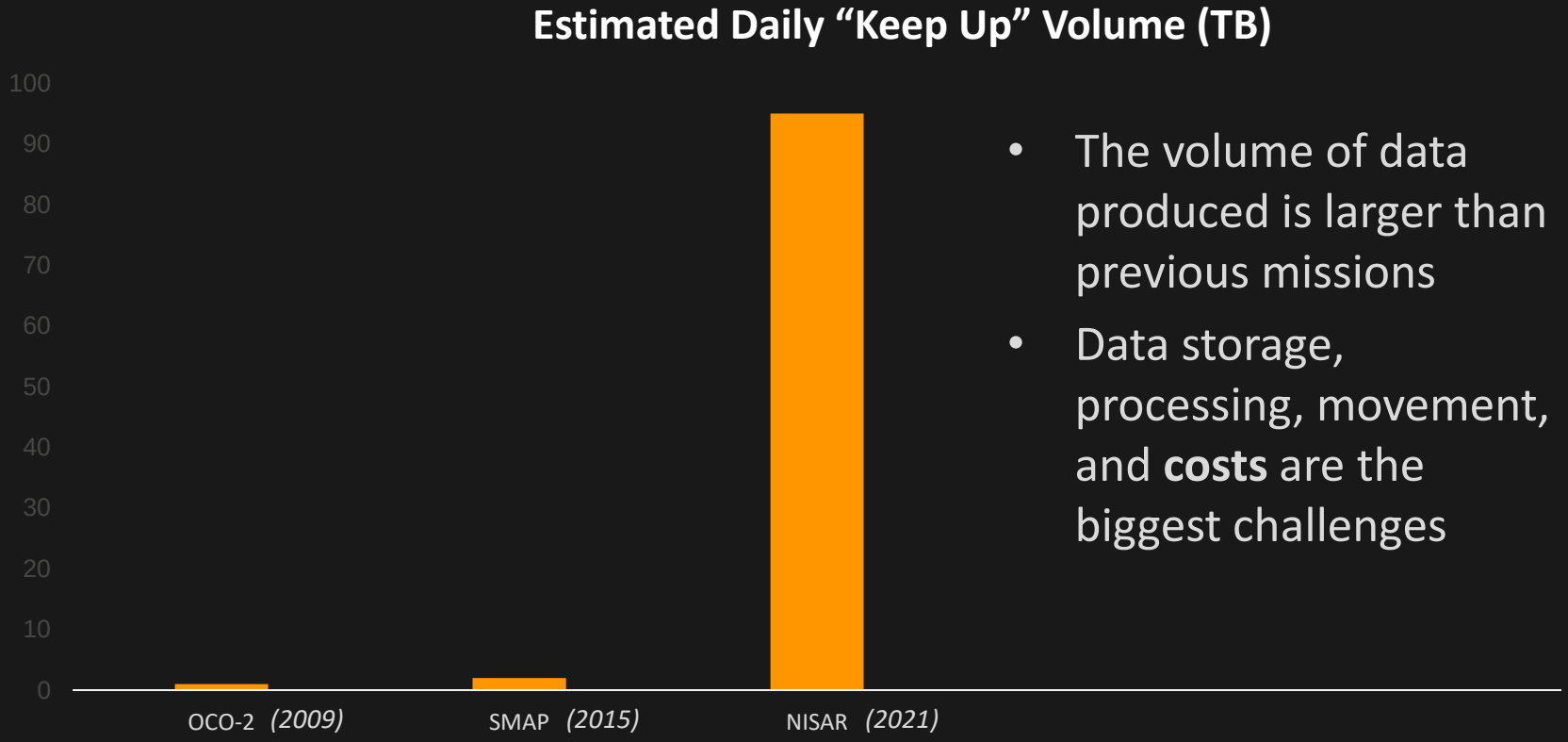
- *“API rate limit exceeded”*

“Jittering” the API calls

- Introduce ***randomizations*** to API calls
- Distributes load on infrastructure



Next generation NASA missions



- The volume of data produced is larger than previous missions
- Data storage, processing, movement, and **costs** are the biggest challenges